COMPUTER ANIMATION WITHIN VIDEOSPACE

by

Gloria Jeanne Brown-Simmons

B.A., Fisk University, 1970

M.F.A., University of California at Los Angeles, 1980

Submitted to the Department of Architecture in partial fulfillment of the requirements of the degree of

MASTER OF SCIENCE IN VISUAL STUDIES

at

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June, 1984

©Gloria Jeanne Brown-Simmons, 1984

The author hereby grants to M.I.T. permission to reproduce and to distribute publicly copies of this thesis document in whole or in part.

Signature of Author . . . .
Gloria Jeanne Brown-Simmons
Department of Architecture
11 May 1984

Certified by . . . .
Otto Piene, Professor of Environmental Art
Thesis Supervisor

Accepted by . . . .
Nicholas Negroponte, Chairman,
Departmental Committee for Graduate Students

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUN 1 1984

LIBRARIES

Rotch

MIT Libraries
Document Services

Room 14-0551
77 Massachusetts Avenue
Cambridge, MA 02139
Ph: 617.253.2800
Email: docs@mit.edu
http://libraries.mit.edu/docs

DISCLAIMER OF QUALITY

Due to the condition of the original material, there are unavoidable flaws in this reproduction. We have made every effort possible to provide you with the best copy available. If you are dissatisfied with this product and find it unusable, please contact Document Services as soon as possible.

Thank you.

The images contained in this document are of the best quality available.

For Malakhi Simmons

COMPUTER ANIMATION WITHIN VIDEOSPACE

by

Gloria Jeanne Brown-Simmons

Submitted to the Department of Architecture on May 18, 1984, in partial fulfillment of the requirements for the Degree of

MASTER OF SCIENCE IN VISUAL STUDIES

ABSTRACT

Scene analysis is explored as a computer animation technique for the purpose of creating a new art form. Tablet drawing routines are used to identify relevant objects by superimposition of lines keyed over the videotaped images. Conventional stereo techniques are used to record image pairs for range calculations in order to provide the animation program with depth data. In this process, the tools used to create the images work in the same domain as human cognition and thought.

Thesis Supervisor: Otto Piene
Title: Professor of Environmental Art

CONTENTS

VIDEOSPACE 5

ILLUSIONARY DEVICES FOR DEPTH PERCEPTION 11

- Drawing and painting 12
- Photography 16
- Cinematography 18
- Animation 19

MOTION CONTROL 22

ROBOTICS 27

PROJECT DESCRIPTION 34

- Recording the site 34
- Test at the AI lab to recover depth 37
- Interactive correspondence 39
- The depth from disparity test 40
- Calculating depth from disparity 41
- Creating the animation 42

STEREO VIDEO TAPES 44

- Marblehead 46
- Boston: Cityscape 47
- Cardboard City 49
- Boston Harbor 51

ACKNOWLEGDEMENTS 54

Sources of Illustrations 55

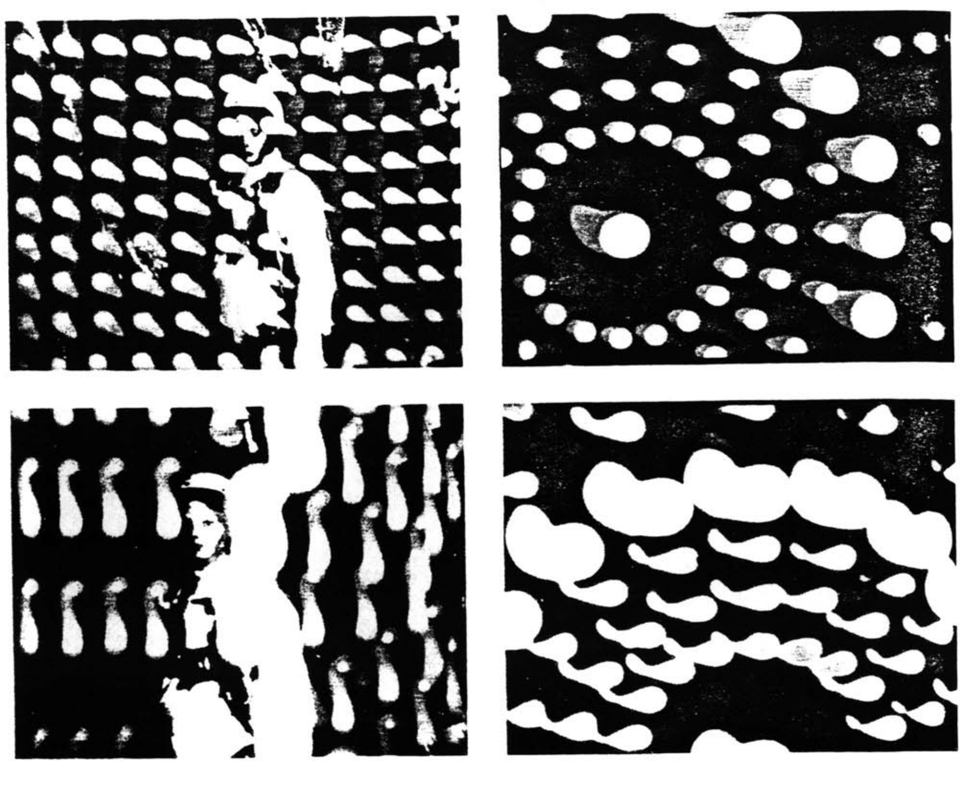
Notes 57

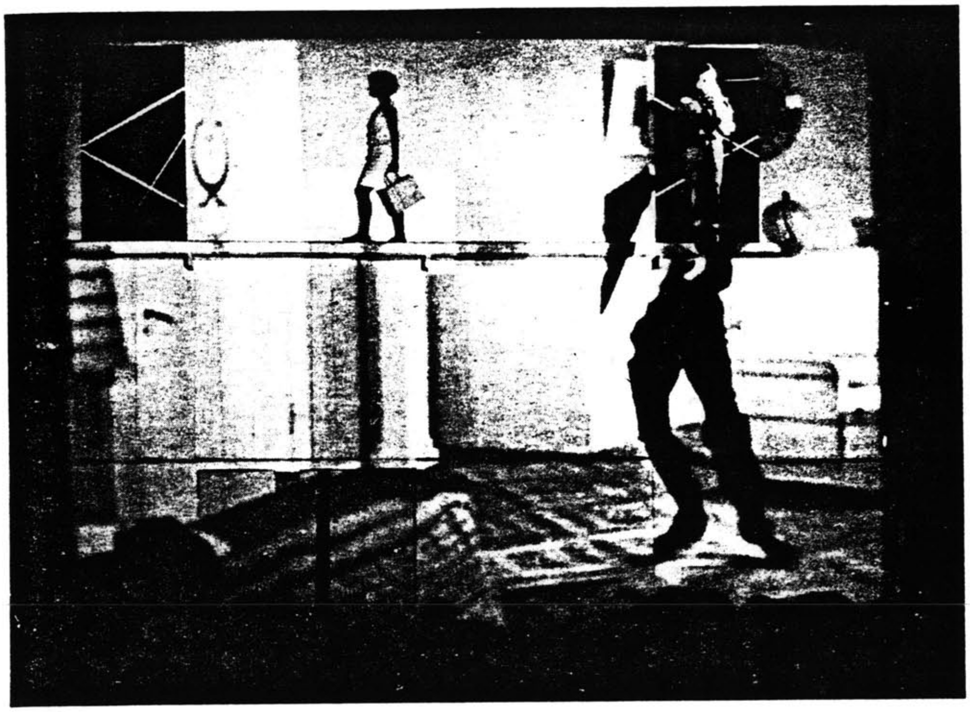
Bibliography 59

VIDEOSPACE

The term videospace is used to describe ways in which electronic artists use video to extend the visual, aural, and temporal experience to aesthetically augment the personal experience. For some artists, the term refers to a sculptural event. The events are called "installations" and are made up of single or multiple monitors arranged and viewed in the same manner as a traditional three dimensional sculpture would be seen in a gallery. For other artists, the term means the space in front of a camera that is "activated" by a viewer performer during a live event. Live (real-time) events are often a combination of prerecorded images and the performer, and are viewed by either the performer, the audience, or both while the event is occuring.

Otto Piene: *Electronic Light Ballet.* 1969.





ONCE Group: *Unmarked Interchange*. 1965. Live performers interact with projection of *Top Hat,* starring Fred Astaire, Ginger Rogers.

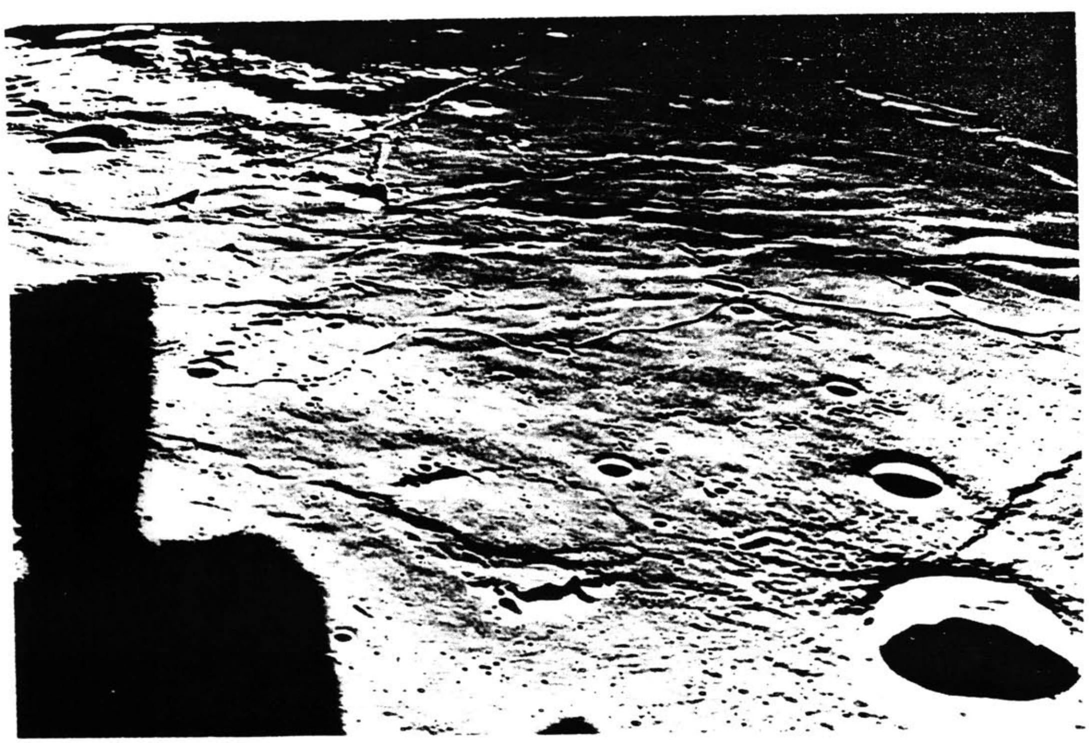
Chroma-Key video matting makes arms of Alwin Nikolais dancers seem to float in space.

Another form of videospace is a cognitive hypothesis that stimulates the viewer to analyze what is taking place visually. The viewer contemplates a variety of visual puzzles constructed from spatial and temporal elements derived from live interactive performance; a traditional reference in film and video for which the space is created by camera movement and perspective;

> 'The camera tracks down the elevator and through the bustling lobby of the Atlantic Hotel as if it were one of the guests there, continues through the revolving doors, and stands with the porter in the rain as he hails a cab. The sequence not only imparts excitment with its movement, by also establishes the size and importance of the hotel...' 1

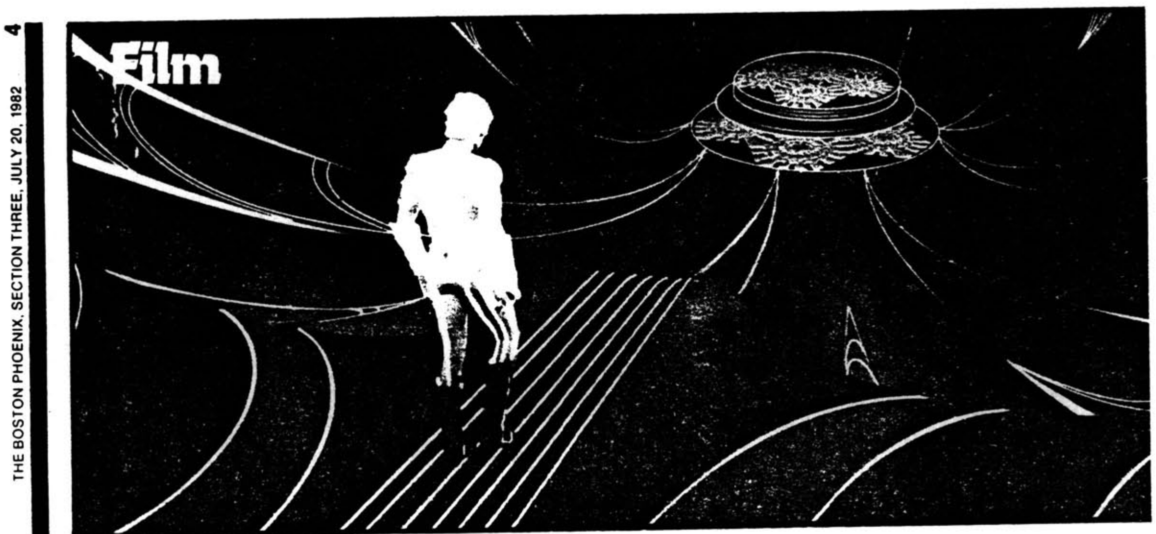
and the space sensed by the viewer of scenes from sites removed from their immediate environment.

> 'The more I work with Video, the more I think about Lessing's distinction of Space art and Time art in the eighteenth century.' 2



Electronic artists are not confined to the real world image in creating videospace. Artists using new imaging techniques visually articulate other types of video space as abstract geometrical forms constructed mathematically, sometimes based on structures found in other mediums such as music or nature.

> 'I look for the medium (computer graphics) to go beyond an imitation of life, beyond fantasy, into universes of form previously unknown, the invisible beauty of mathematics and the dance of physics will become visible...' 3



Lost in the TRONhouse

Videospace, as refered to in this thesis, is defined as a real world image and its mathematical (three dimensional) model. This analytical videospace is similiar in aesthetic to the real time insertion of performers (persons or objects) over an image, but differs in concept with the addition of control over the three dimensional spatial arrangement.

In the comtemporary works of artists who mix separate images i.e. Paik and Dewitt, the relationships in time, form, color, and line are well developed. These compositional elements have precendence over the creation of a spatial relationship between the layers of assembled images due to the technical limitations of video keying.

> 'As collage technic (sic) replaced oil paint,
> the cathode ray tube will replace the canvas.' 4

The lack of control over a three dimensional representation of the image necessitates reliance on other illusionary techniques for producing a sense of space; i.e. lighting, scale, and occlusion. But this illusion is usually destroyed as the separate images are introduced into the scene or moved around within the scene, thus appearing as flat cardboard-like shapes.

As opposed to the casual mixing of multiple two dimensional images, the construction of a scene model affords the artist a higher level of interaction and manipulation of the actual space. In structuring the match between real experience and generated images, the existing internal image (imagination) is reconciled with external (selectively recorded) events. The external events provide material for internal improvisation.

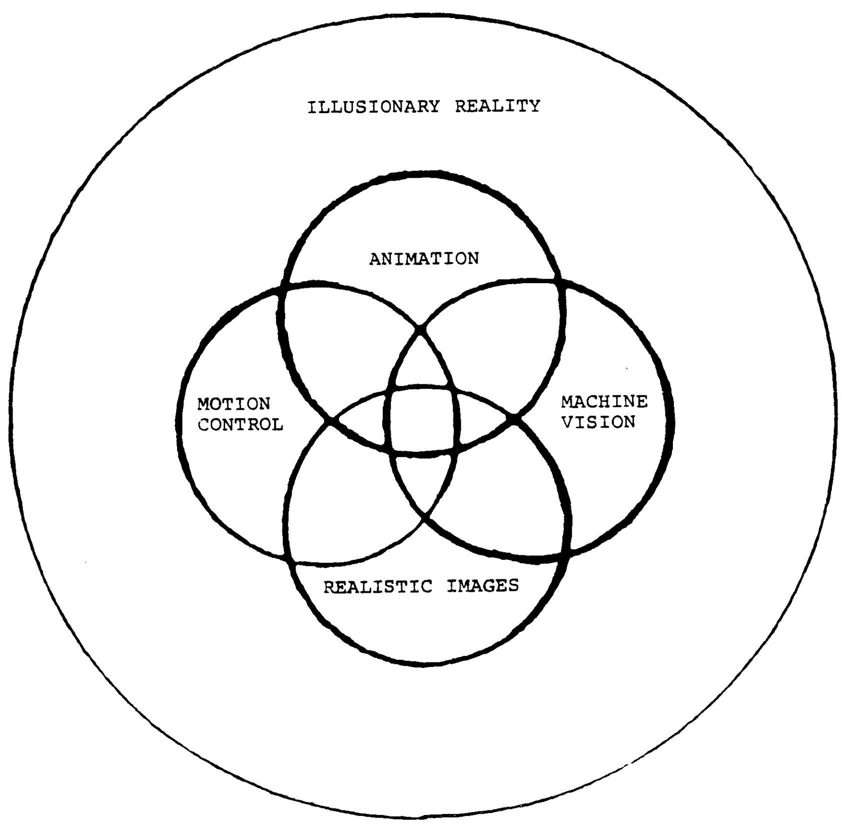
Ultimately, the process of "forming real-space" becomes invisible. The state of "willing suspension of disbelief" is induced in the artist during the production of the art to remove issues of mechanics vs. image. Thus, it follows that with the perfection of the image the viewer will be convinced of the actuality of the art.

> 'Far away, at the very limit of distance, lay range upon range of snow-peaks, festooned with glaciers, and floating, in appearance, upon vast levels of cloud... There was something raw and monstrous about those uncompromising ice cliffs, and a certain sublime impertinence in approaching them thus. He pondered, envisioning maps, calculating distances, estimating times and speeds.' 5

Peter Campus: *Shadow Projection*. 1974.

ILLUSION DEVICES FOR SPACE PERCEPTION

> 'For in man's evolution, images antedate words and thought, thus reaching deeper, older, more basic layers of the self. Man begins with what he sees, progressing to visual representations of reality. Their transmutation into art does not seem to diminish the images' impact. As holy today as in man's prehistory, the image is accepted as if it were life, reality, truth. It is accepted on a feeling - rather than mind level. Significantly, it is only if the 'suspension of disbelief' is broken by dissatisfaction with a given film that the viewer emerges from his hypnotized state.' 6



In early cave paintings, animals were modelled in ocher and often painted over a protruding portion of the wall to emphasize the volume of the animal. Several animals were sometimes drawn overlapping one another, employing occlusion as a technique to denote distance from the viewer.

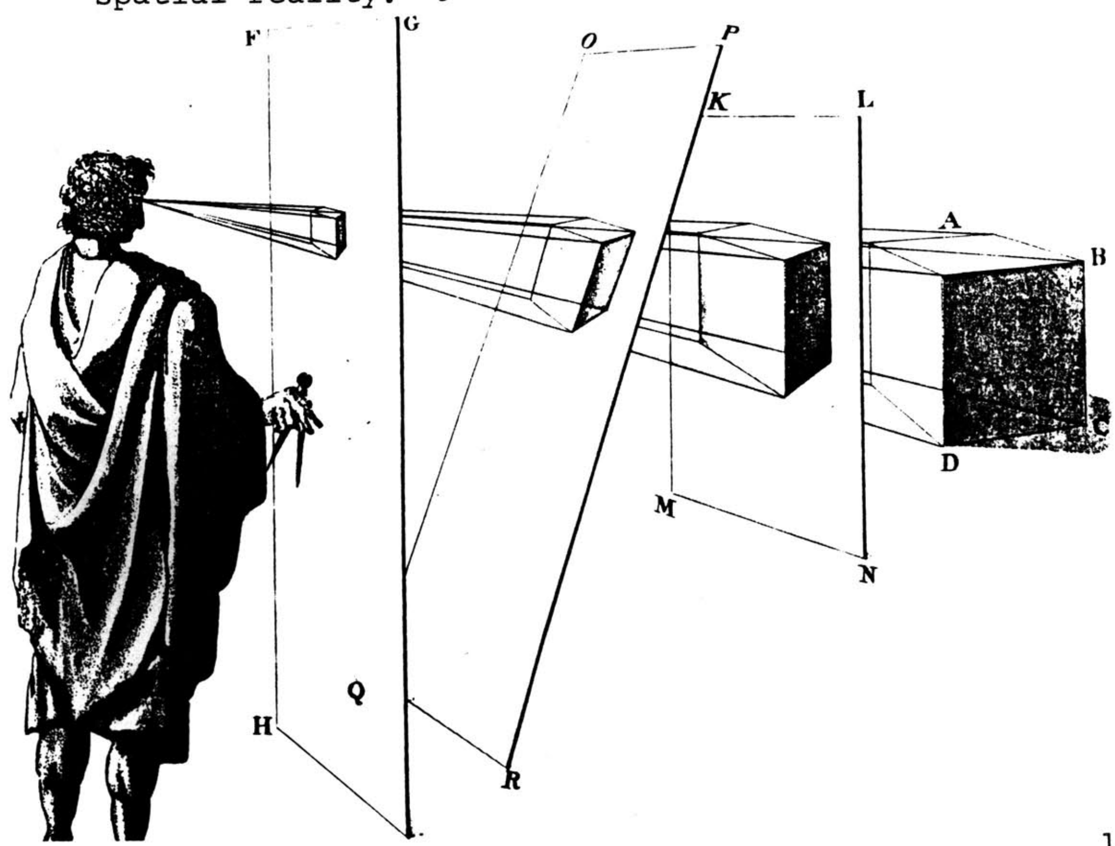
> 'If one spatial form obstructs our view of another form, we do not assume that the second ceases to exist because it is hidden. We recognize, as we look at such overlapping figures, that the first or uppermost has two spatial meanings - itself and beneath itself. The figure which intercepts the visible surface of another figure is perceived as nearer.' 7

Occlusion, scale, and occasionaly height, denote a sense of spatial distance in ancient art.

' Structure is the anatomy of space, but it also represents the total investment of the artist; the vehicle of his means, and the filter of his inspiration and imagination. In the end, in a completely realized work, space and structure become one.' 8

Historically, attempts to create or reproduce realistic spatial relationships, for the purpose of illusionary reality, have evolved with the development of the perspective system. This system was first conceived as a cone of vision (Agathaclus, Anaxagoras, and Democritus). Later, the cone was interpreted as visual rays emanating from the eye to an object (Euclid, Archimedes, and Apollonius).

'As the Eucidean geometry was but a first approximation in the knowledge of spatial forms, reflecting only a certain limited complex of spatial properties, the traditional forms of visual representation were but the first approximation in sensing the spatial reality.' 9



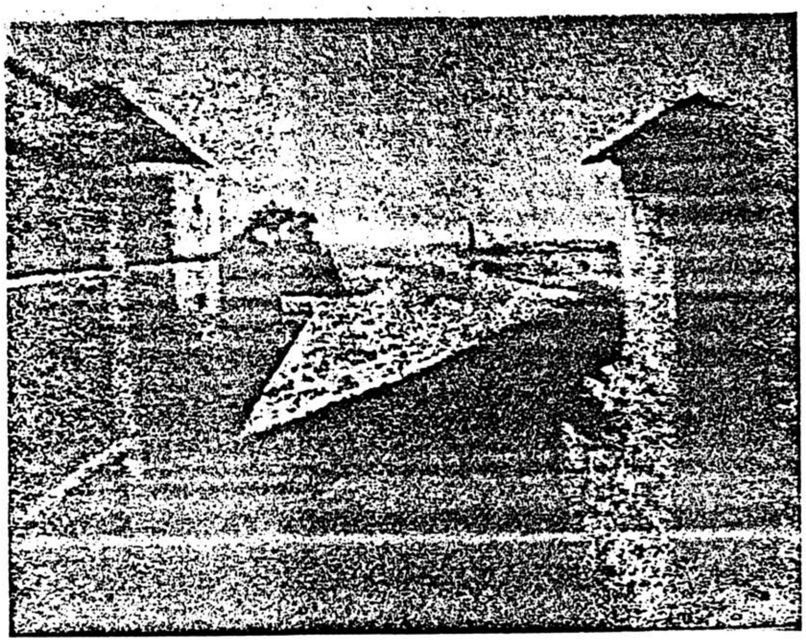
Vitruvius described the perspective system as a drawing in which all lines meet at the center of a circle in DE ARCHITECTURA. The paintings' depth was enhanced by the addition of projections, foreshortening, and multiple vanishing points during the late 13th and early 14th centuries (Duccio and Giotto).

A perspective system based on mathematical geometry was developed by Brunelleschi. Alberti defined a structure for painting as a cross section of the visual pyramid known as the focus system, by which an artist constructs the painting on a perspective grid. Further developments lead Piero and Leonardo to extend Alberti's system to include a method for constructing perspective from top and front views. About the same time as Leonardo, Durer wrote his treatise describing both mathematical and mechanical perspective. Desargues was to lay the foundation for projective geometry in his treatise on conic sections. Monge applied the multiview orthographic projections developed by Piero in engineering drawing and is considered the initiator of descriptive geometry.

> 'Monge developed his principles for the geometrical solution of spatial problems in 1765 while working as a draftsman of military fortifications. His solutions were at first regarded with desbelief, but were later guarded as a French Military secret.' 10

Among Monge's students were Poncelet, who revived projective geometry; and Crozedt, who introduced Monge's principles to the United States at West Point. He was followed at West

Point by Davies, who published the first U. S. treatise on descriptive geometry; and Church (Davie's successor) who published ELEMENTS OF DESCRIPTIVE GEOMETRY, in 1864, which became the leading U. S. textbook on the subject through to the first decade of the 20th century.



9 Nicéphore Niepce. The first photograph, 1826

With the invention of the camera, representational painting receded in importance and other schools soon developed. These new schools rejected the single view perspective as the unifying element of a picture. The Cubist painters experimented with multiple perspective views; others returned to the spatial primitives of color, line, form and two-dimensionality, developing each element into entire themes.

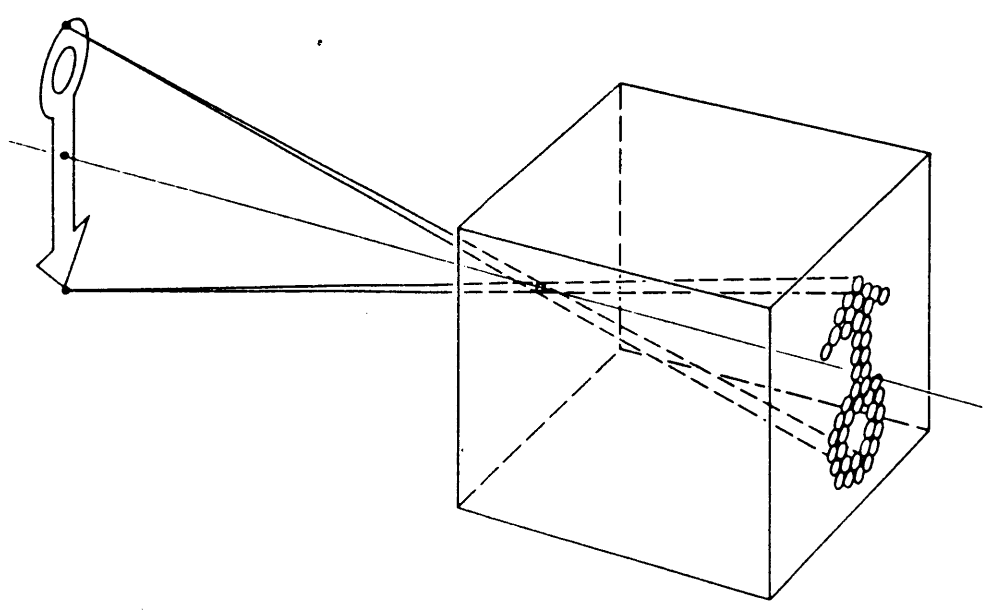
> 'Everything we call nature, is, in the last analysis a fantasy picture." Malevich says, "with not the least resemblance to reality." 11

> 'The underlying philosophy of this final purification of the picture-plane from the object world led to ultimate rejection of any attempt to represent objective reality.' 12

Contemporary artists are becoming interested in illusionary space and also mixtures of flat space and three-dimensional space.

PHOTOGRAPHY

Depth clues for monocular photographs are: linear and size perspective, interposition, shading and shadow, texture gradient, aerial gradient, and gradient of details. Early cameras were based on the "camera obscura" model developed by Renaissance artists for projecting scenes onto walls for mural painting. The first camera used film that required long exposure times. With the development of smaller cameras and faster exposure times, the photographer was able to compose shots with the image plane in various positions.

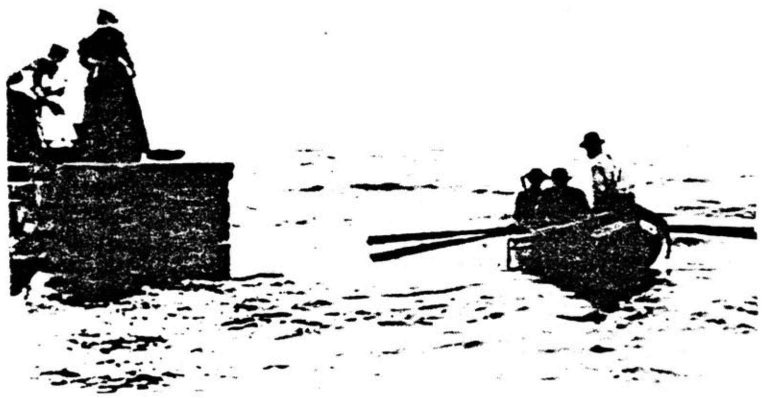


As opposed to the convention of exposing the image plane parallel to the principle plane of the object, freedom to move the camera provided new "perspectives" on traditional scenes, i.e. a bird's eye view, a frog's eye view.

> 'Amplified perspective is used in photography, photomontage, and in the motion picture as a potent device for creating a strong sense of space.' 13

Amplified perspective is accomplished by a combination of lens characteristics and composition. The dramatic qualities of space can also be rendered with contrast and shading the surfaces with light. Texture and detail are magnified and emphasized in realistic photographs; surfaces become less defined as they appear to recede.

The freedom of the outdoors and the excitement of motion— Lumières' BOAT LEAVING THE PORT



CINEMATOGRAPHY

Cinematography maintains the same depth clues as photography with the addition of a greater sense of place by time of movement within the space and additional spatial information from various shots (long, medium, close). The camera plays an important role in developing the space as opposed to simply recording a portion of it. Sound to image relationships infer distance as well.

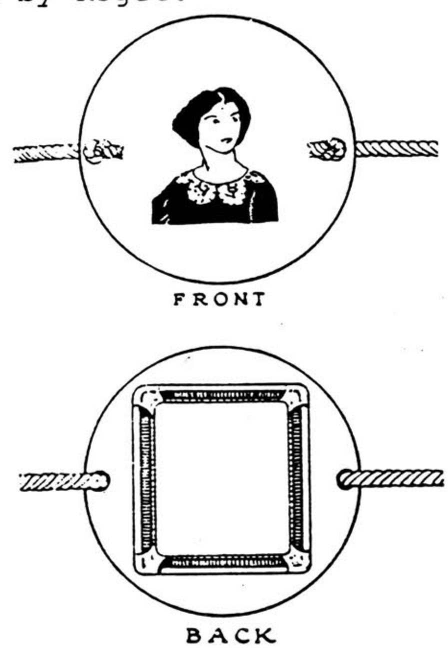
The 'reality' of the flat, two dimensional screen – secret repository of our deepest dreams – is here revelled in by a love-struck man who projects moving images of his beloved on to his body and attempts to caress them; but films come to an end and there is pathos in his unrequited, twice removed passion. (Roberto Rosselini, Virginity, *episode in omnibus film,* Rogopag, *Italy, 1962)*

ANIMATION

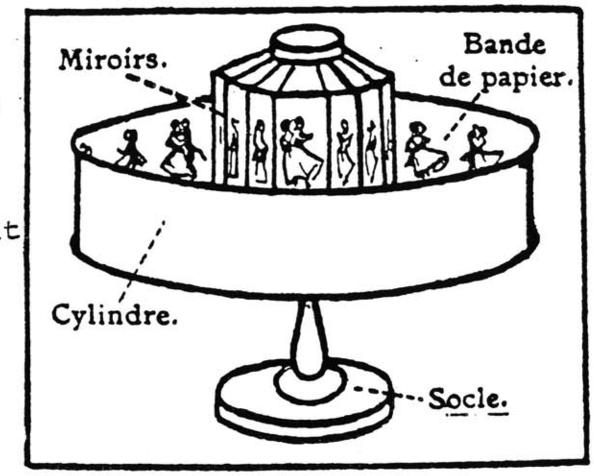
Animation is a series of drawings, paintings or photographs, in which the elements are transformed, so as when projected over time, appear to move.

The earliest known animations were optical toys. These toys were being produced in France, England, Germany and the United States, and they all used drawings for their images. One such toy was the Thaumatrope, credited to Dr. Finton, in 1825 (also credited to Dr. Paris for the same year). The Thaumatrope consisted of a cardboard disc with pictures on both sides. When the disc spun around, the pictures appeared to merge together. The optical toys were based on persistence of vision which, although known by the ancients, was being investigated during the early 1800's, and written about by Roget.

Plateau published his investigations on persistence of vision in 1829, and patented the Phenakistiscope in 1832. The Phenakistiscope was similiar to the Zoetrope designed by Horner in 1834. Uchatius was noted as being the first to project animation by light. By running



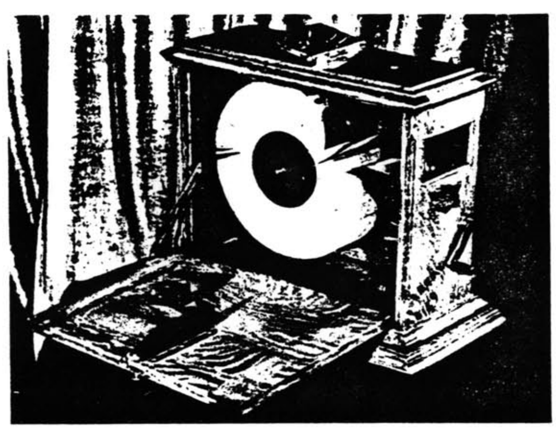
with a torch behind a series of projectors, he cast images on the wall. Each projector had a slide with an image that was slightly displaced from the previous one. When the slides were projected in rapid succession, superimposed on the same spot, the image appeared to move. By 1853 Uchatius had developed the Projecting Phenakistiscope. With the invention of film (Niepce 1816) to capture real pictures, the photographic process was soon applied to animation.



'The first attempts at motion photography were posed stills that simulated continuous action. The stills were projected with a Projecting Phenakistiscope to give the appearance of motion.' 14

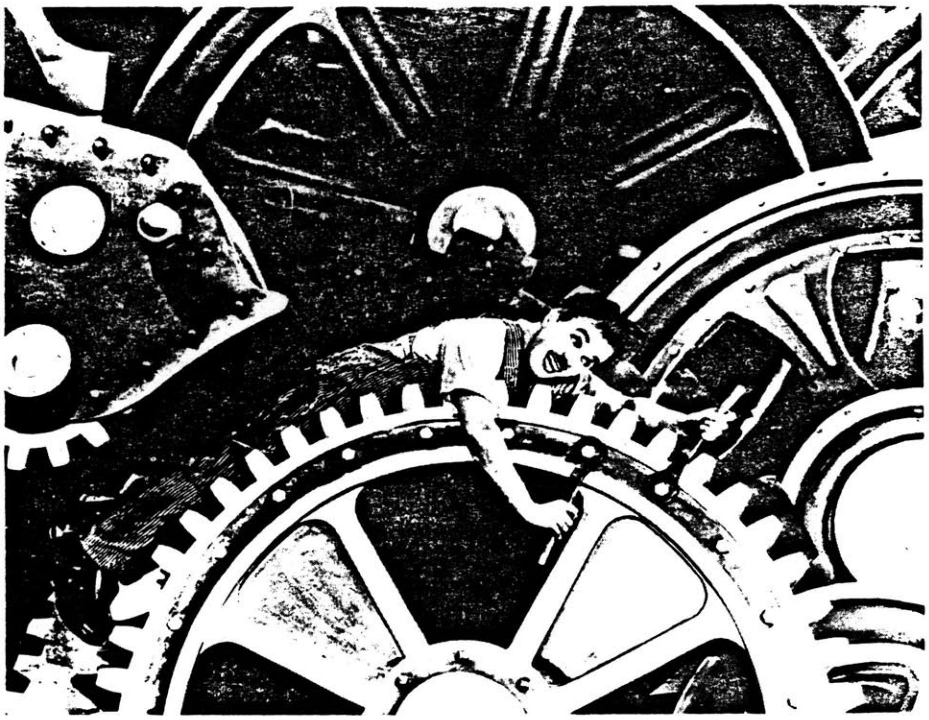
Mutoscope viewing machine (*Collection of Kemp R. Niver and J. P. Niver*)

Flip-card book (*Collection of Kemp R. Niver and J. P. Niver*)



The impact of the film process on the artistic community lead them, in addition to non-representational forms, to incorporate time as a compositional element. As Picasso said "I paint what I think, not what I see" so have the abstract filmmakers (of which Picasso was one) felt the need to refine their methods of production to realize their imaginations. In their early works, artists were finding ways of expressing their new ideas for visual/musical structures. The painter Survage studied movement by painting on glass (1912-1914) each glass plate a "key frame" in a sequence of abstract animation. Walter Ruttman turned his paintings into motion studies based on musical structure; Viking Eggeling and Hans Richter painted "scrolls" (graphical music) and produced films that 'discovered the importance of an esthetic in film' (Moholy-Nagy). Students at the Bauhaus began to experiment with "Reflected Light Compositions", 1922, by moving colored templates back and forth in front of a spotlight. "Reflected Light Compositions" expressed the age-old longing for making the immaterial visible, for the picture sublimed by dematerialization. MacLaren, Lye, Fischinger were to follow in the succession of abstract animators.

Optical illusions, foreground to background reversal, and use of a perspective system that contained the abstract shapes (implied by scaling) provided the spatial context for these films.

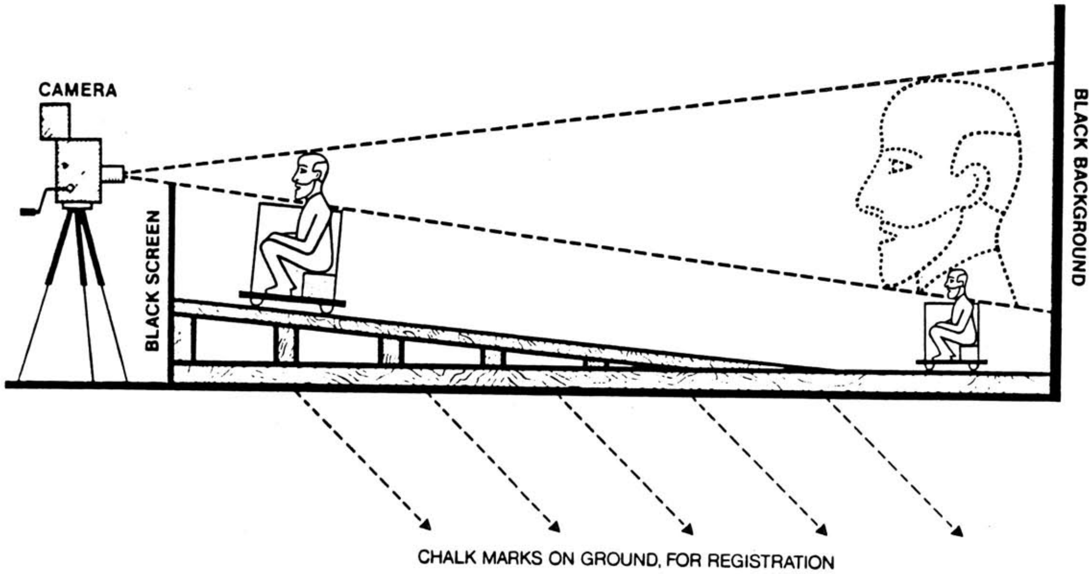


Modern Times.

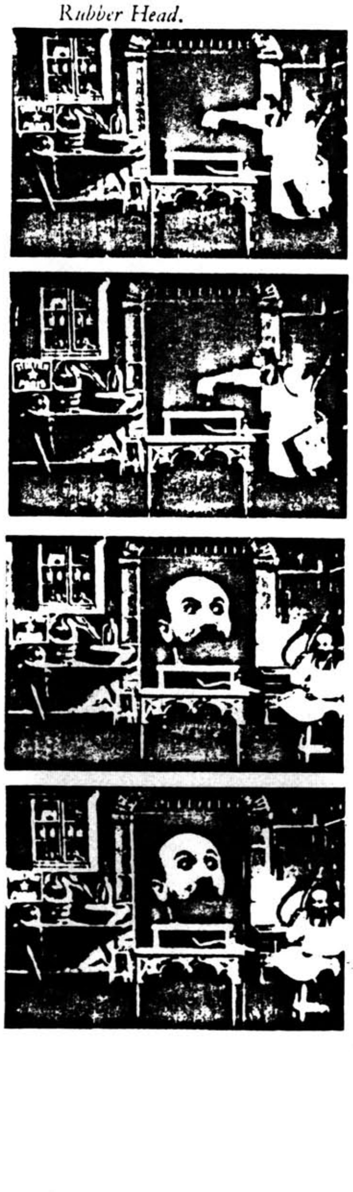
MOTION CONTROL

Montion control, in cinematography and videography, is any apparatus that will move through a set of points on a predetermined path. The simplest system for motion control is the camera and camera operator. According to how precisely the camera operator can make a move determines the accuracy of the system.

Motion control devices increase in complexity by the number of repeatable moves, directions, and complexity of direction; i.e. zoom the camera from 25mm to 50mm in 3 seconds, do while panning from 10 degrees in the x direction to 50 degrees, and tilt at a rate of -2 degrees per second in the y direction, repeat this sequence 200 times (some systems can repeat up to 500 passes).



How Méliès' head grows in *The Man With the Rubber Head.*

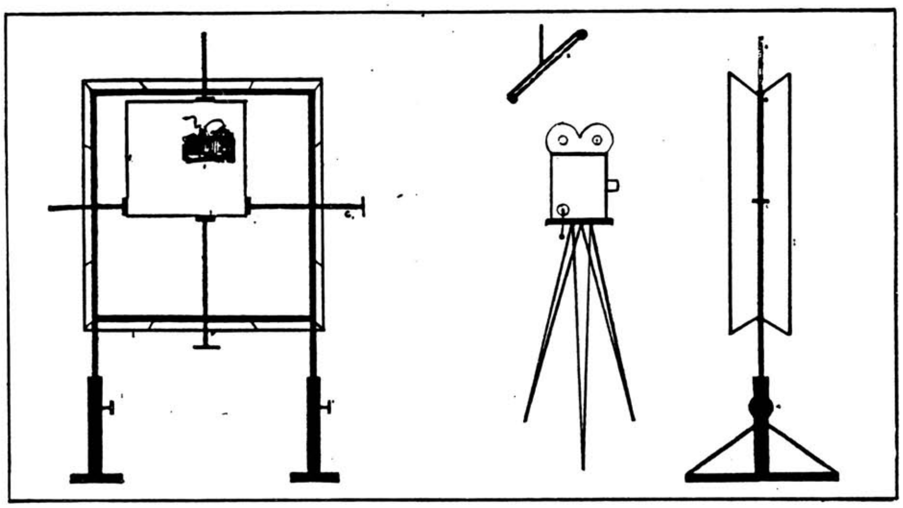


The conjurer, Georges Melies (1861-1938) used the camera as a way of capturing his illusions for theatrical distribution and as a technique to heighten the "realism" of his magic. His first camera, the Kinetographe, 1896, was based on the Cinematographe (Lumiere 1895). His techniques for producing special effects for films such as THE MAN WITH THE RUBBER HEAD (1902) made use of a mechanical motion control system. Melies

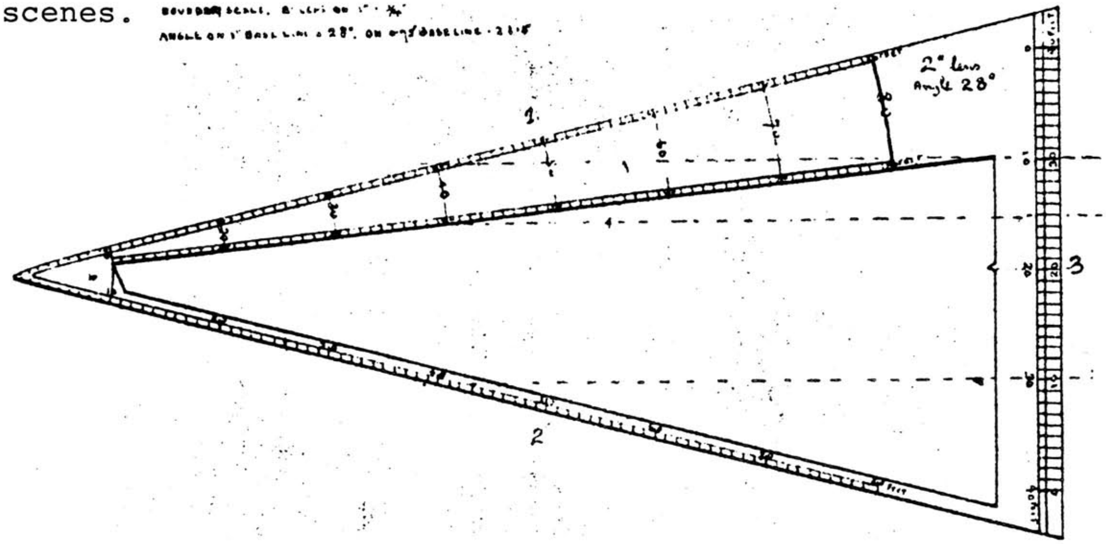
> 'sat on a dolly that was drawn towards the camera, a screen hiding his body and the machinery.' 15

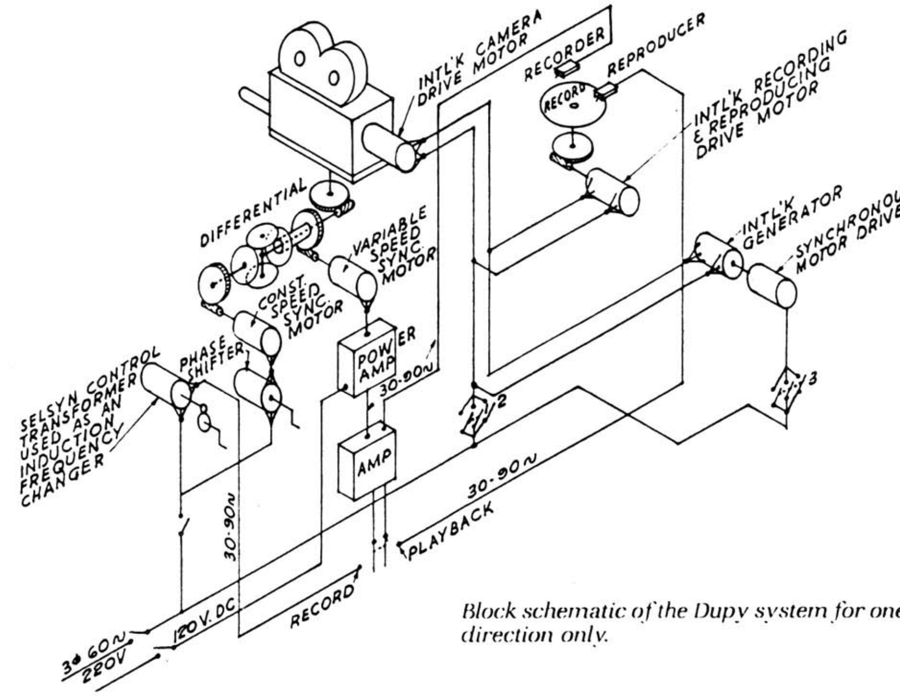
Melies clearly saw the film as stageplay, and he referred to his technique as making "artificially arranged scenes".

After the turn of the century, there was interest in accomplishing insertions and double exposures filmicly. James Brautigam developed a mechanical repeating motion system for the camera, producing double exposures while working for Thomas Edison in 1914. A method for inserting miniatures and models into "real-life" scenes was described by Alfred Hitchins in 1922.



There were charts made for calculating the field of view in order to insert scaled props in front of background scenes.





Block schematic of the Dupy system for one direction only.

The Dupy Duplicator and the Jennings/Stancliff repeater systems were developed during the forties Both systems were able to pan and tilt the camera, repeating this movement exactly for each element of the shot (frame accurate); both used an audio track for recording and playback of the camera.

John Whitney started the first motion control company. His system combined analogue computers with an animation camera. The slit scan and streak effects were two of many created.

> 'Trumbull built a slit-scan machine. His was slightly different from John Whitney's machine in that it was arranged horizontally rather than vertically and he borrowed from scientific and industrial techniques.' 16

John Gail developed a system for combining 'real' actors with miniature sets in real-time. The system became the Magicam and was later moved to Paramount where he made COSMOS CALENDAR. The system was capable of combining the two images maintaining

> 'a real time perspective match between scene components under dynamic conditions. This unit in essence simulates a six degree of freedom pantograph coupling two cameras at different scene scale factors.' 17

The control over the six degrees of freedom has made the Magicam system attractive to robotic groups such as Boeing Aircraft were the system is used in their 767 simulators. With the development of similiar, sophisticated, motion control systems the field has intersected the robotics industry.

> 'The ability to repeat a motion has made it possible to integrate the effects shot into live action in such a way that audiences accept the impossible as probable and the dazzling as typical.' 18

By varying the visual palette (views), the audience can better comprehend the spatial relationships of the objects in the scene. With the addition of control over the time and motion between views, the self-to-object relationship is accentuated and further supports the perception of the scene's elements. Furthermore, the addition of detail, by selective exposure, to scene components enhances the the texture gradient.

ROBOTICS

Early robotics was developed prior to World War II with primitive teleoperators and numerically controlled machine tools. The teleoperator was a system by which an operator could perform some task at a distance (remember those games at the amusement park where you had to operate an arm to pick up the prize?). The numerically controlled machine had a set of gears that performed a task as recorded on punched paper tapes.

Wiener's CYBERNETICS: OR CONTROL AND COMMUNICATION IN THE ANIMAL AND THE MACHINE (1948) was the first book on the theories of

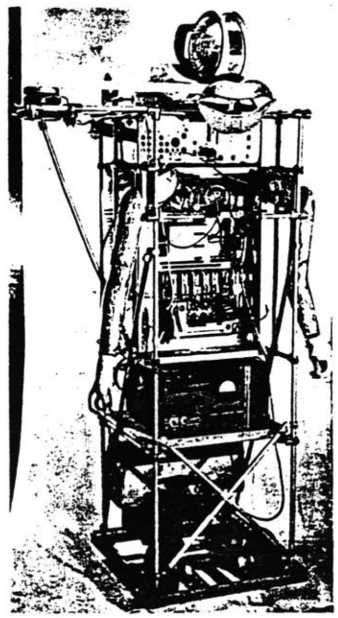
> 'regulated functions of the body and the protocol of electromechanical control.' 19

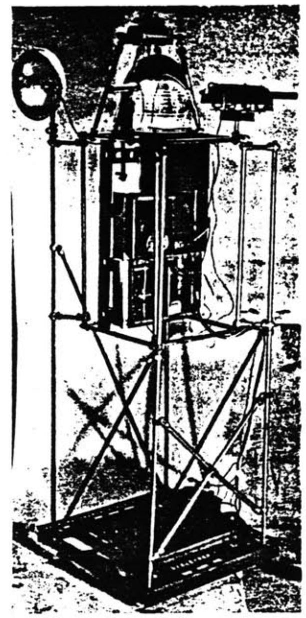
He, with Bigelow and Rosenbluth, pioneered what is known today as robotics - that is machines that are patterned after human beings. Wiener's works were widely read and had much influence on the artistic community. Although the means to create truly human machines were beyond the scope of most artistic projects, some artists did mimic the concepts of man/machine construction.

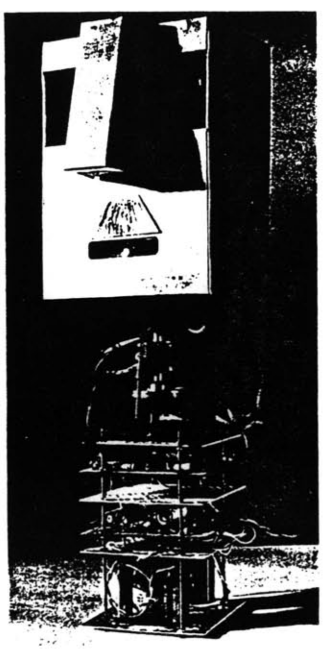
> 'The historian proceeds to link Western culture with an unstoppable craving to wrest the secrets of natural order from God - with the unconscious aim of controlling human destiny, if not in fact becoming God itself. The machine, of course is the key to this transference of power. If it constructs our destiny, it can do no less than become the medium through which our art is realized.' 20

Right, *Rosa Bosom (R.O.S.A.–Radio Operated Simulated Actress)*, 1965
Electronic parts, aluminium, batteries, motors, etc.
24 × 36 × 72 in.
Originally designed as an actress to integrate with live performers in production of *Three Musketeers*, at the Arts Theatre 1966, where it played the Queen of France.
Method of control: radio; programmed; environmental.

Far right, *Mate*, 1967
Electronic parts, aluminium, batteries, motors, etc.
24 × 24 × 72 in.
Built as companion to R.O.S.A., which it follows automatically and generally interacts with. Using ultrasonics, infra-red and sonic signals.

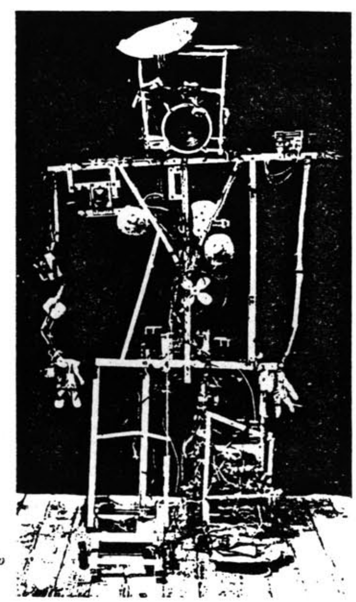






Albert 1967 by John Billingsley

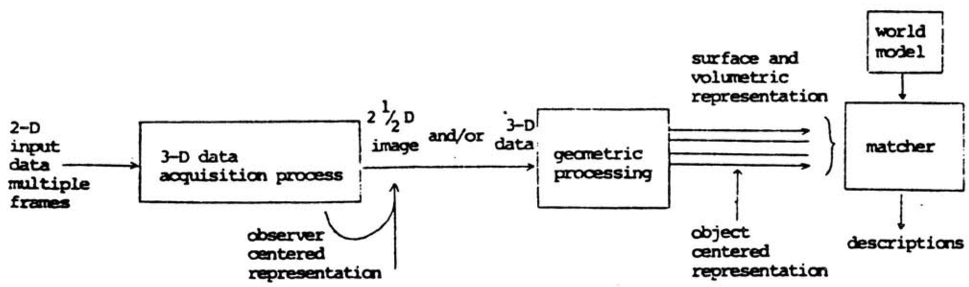
The sensing mechanism uses modulated light to obtain a 'turn left' or 'turn right' signal. The eyes are made to flicker in antiphase, and the reflected light is picked up by a photocell in the mouth. The signal is then amplified, filtered and demodulated, and used to drive the motor. If more 'left hand' light than 'right hand' light is seen by the photocell, the motor turns the head to the left.



Nam June Paik, *Robot-K456 with 20-Channel Radio Control and 10-Channel Data Recorder*, 1965.

The field of cybernetics developed beyond the study of electromechanical similarities between humans and machines to encompass the realm of thought - Artificial Intelligence.

A branch of Artificial Intelligence, Scene Analysis, can be used to extract conceptual data from images. Machine vision techniques are used to process images visually before attempting to understand them conceptually. Methods used in these fields follow experimental models for human vision and cognition, attempting to parallel the image understanding process in the human visual system.

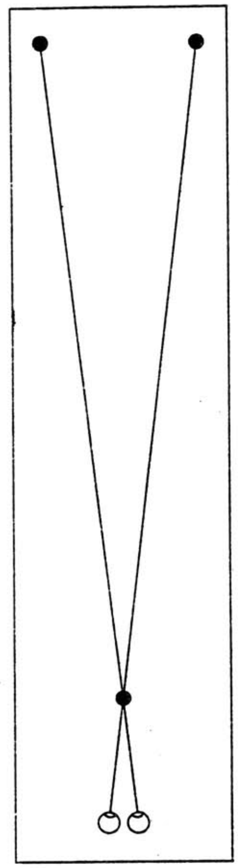


A STRUCTURED DIAGRAM OF SCENE ANALYSIS

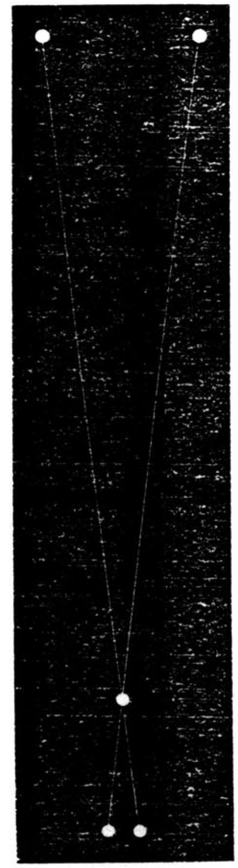
Tools developed in this field can be particularly useful to artists who work with images. Specifically this artist used a system to extract three dimensional data from stereo images. Artificial computer generated images were rendered in conceptual three dimensional space - a basic level of human cognition. The images were then combined with the live video from which the three dimensional space was originally derived. The image creation process uses techniques that are based on models of the cognition of the image.

The symmetry in this process means that the tools used to create the image are tools that work in the same domain as human cognition and thought.

There have been works created by artists that tend towards this concept. Using stereo imaging techniques, they have simulated their interpretation of the human vision system. For example, Alfhons Schilling, in his THE EYES, has created a binocular video system for synthesising a completely subjective environment. The viewer wears individual eye monitors while Schilling composes the images with hand-held cameras.



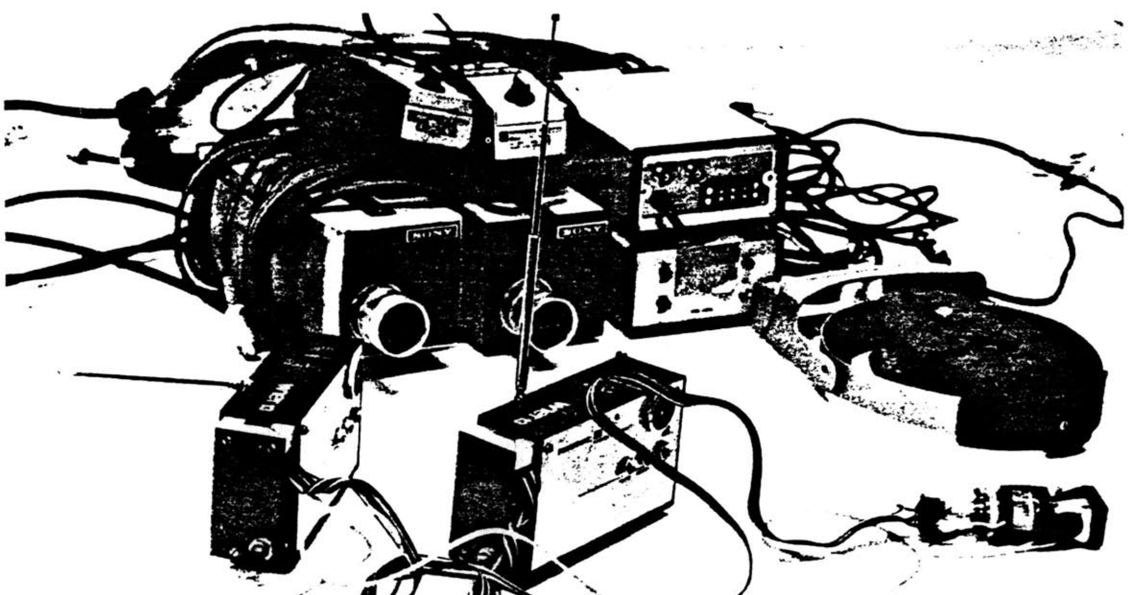
The eyes signal the distance of near objects in very much the same way as the astronomer's parallax method for stars. In both cases, these are *direct* methods for determining distance and can be *unambiguous*, if the base-line is known.



Astronomers measure the distance of stars by their parallax shift across the curtain of very distant 'fixed' stars. The astronomer must know the length of the base (twice the distance of the sun) or all his measures will be systematically in error.

The introduction of an 'imaginary' object by stereo methods was produced by Ivan Sutherland in a project that combined a computer model 'over' a view of a room (live, unrecorded background). This thesis goes a step further by providing the ability to record any background and fit the graphics to the scenes geometry.

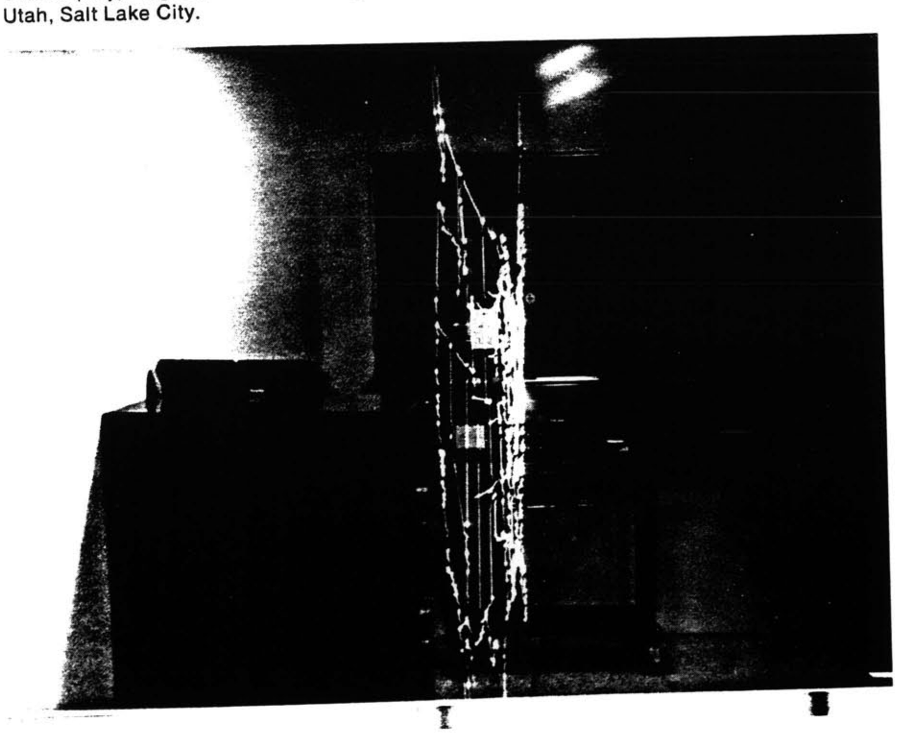
Binocular video system for synthesising a completely subjective environment


SONY
SONY

As scene analysis techniques become more powerful artists will work with images at conceptual levels of greater complexity. Modelling not only the three dimensions that the user will perceive but other conceptual aspects of the image as well.

> 'Thus, behind much art extending through the Western tradition exists a yearning to break down the psychic and physical barriers between art and living reality, not only to make an art form that is believably real, but to go beyond and to furnish (images) capable of intelligent dialogue with their creators'.' 21

Ivan Sutherland. Map of the United States as seen by an observer wearing the head-unit display, August, 1970. Photo by Computer Science Communications, University of Utah, Salt Lake City.



PROJECT DESCRIPTION

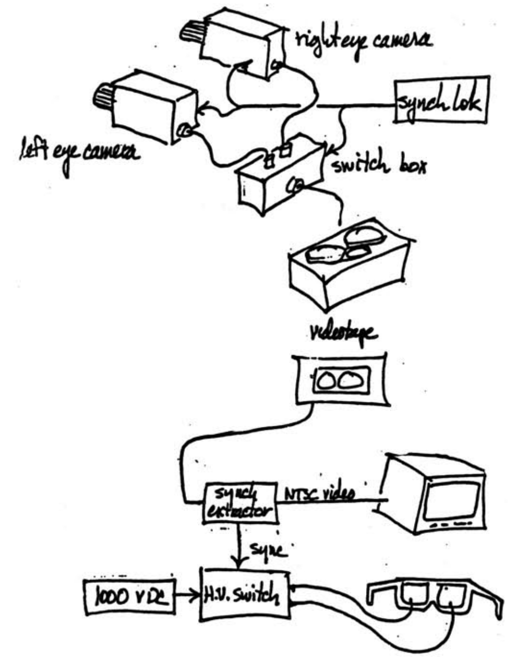
RECORDING THE SITE

The main problem when recording a site to use as a computer model is capturing the three-dimensional coordinates. The use of a motion control system as described in a previous section increased the amount of equipment needed for portable use and also proved to be too costly for the average independent artist to obtain. Photographing the site in stereo, and later, during the post production, using computational stereo methods for scene analysis reduced the field equipment to two video cameras, two microphones, the video tape recorder, NTSC monitor, switching unit and the electronic goggles.

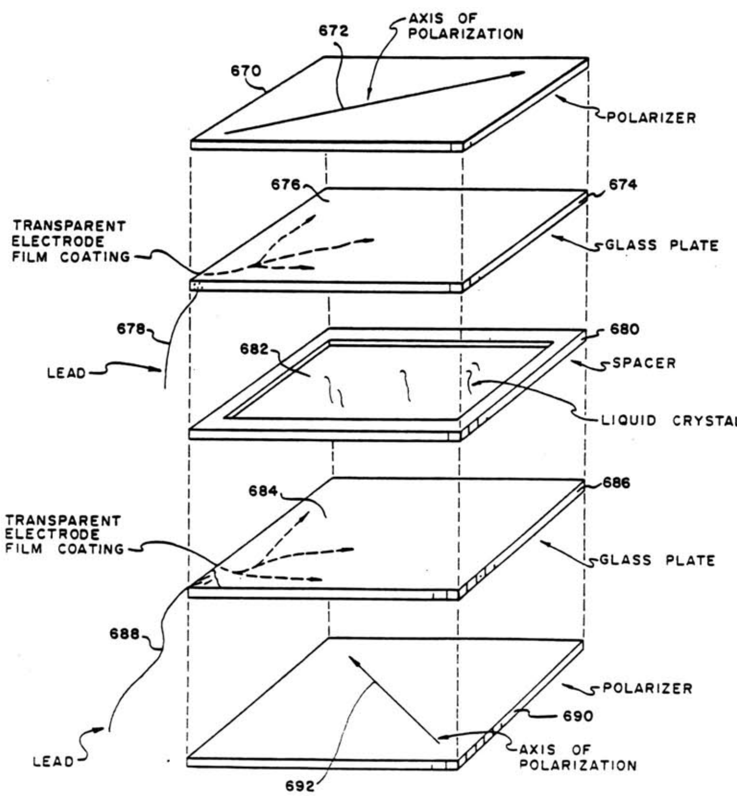
The cameras were JVC, model S-100U, 1" SATICON tube, color, with a resolution in the horizontal direction of 670 lines (calibrated) and a resolution in the vertical direction of 504, 2:1 interlaced, producing a standard NTSC signal.

The switching unit was the stereo viewer system produced by the Ceramics Center at Honeywell based on joint research between Honeywell and the Naval Undersea Center, and patents by John Roese for a liquid crystal stereoscopic television system.

The basic system configuration for producing stereoscopic video is shown in the following diagrams. In this system, shutter goggle stereo TV system



LIQUID CRYSTAL
ELECTRO-OPTIC SHUTTER



two television cameras are operated synchronously to produce a time sequence of alternating odd and even field scans. The function of the raster scan switching module is to switch between the outputs of the two cameras at the field scan rate of the system (the NTSC field scan rate is 1/60th of a second). The output of the raster scan switching module is a single channel, normal, bandwidth video signal, consisting of an interlaced sequence of even field scans from one television camera and odd fields from the other camera. This system is also well suited to video tape recording since the video bandwidth requirements for alternating field stereoscopic images are identical to those of conventional video systems. The video tape recorder was the portable SONY VO-4800, 3/4" U-MATIC deck (resolution of 250 lines for color).

A monitor with a large display screen was needed in order to check the vertical and horizontal alignment of the two cameras. The weight and size of the monitor was inhibiting during the site recording. For large productions, a portable unit with remote cameras would provide more flexibility.

The cameras were mounted on two separate tripod heads; the heads mounted on a dual camera bar which could be adjusted from four to twelve inches (distance between cameras); the bar was mounted on a third tripod head. The three heads had bubble levels so that the mounting components could be adjusted parallel to the ground plane. The two heads for the cameras also had calibration marks inscribed on the pan and tilt components so that the position of each camera could be adjusted independently. The mounting unit could be removed fron the ground base with a quick release enabling the user to change shooting positions. The fluid head provided smooth pans and tilts while the relationship between the two cameras remanined fixed.

The PLZT goggles were used when checking the aesthetic composition of the (stereo) scene. An optically cemented sandwich of antireflective coated glass, a front polarizer; the PLZT wafer of lead, lanthanum, zirconate and titanate ceramics; and a rear polarizer comprised the lenses. The shutters switch at the monitor's field refresh rate. For each frame, the perspective view for one eye is seen during the first field scan, while the other eye's view is blocked. This process is reversed for the second field scan to accommodate the perspective view for the other eye. Repetition of this sequence at normal television frame rates causes the observer to merge the perspective views for both eyes into a single image with a well-defined depth of field.

TEST AT THE AI LAB TO RECOVER DEPTH

In a test to recover depth from the stereovideo, a black and white tape was produced with one tube NEWVICONS of a

Boston Cityscape. The images from this tape were digitized and the left and right interlaced views were sorted into separate images. Using a pixel level boundary detection heuristic (a type of image processing program), the two images were then partitioned into regions. A pattern matching heuristic was then used to mark corresponding edges (and associated points) in the two views from which a disparity could be determined. From the disparity and parameters, of the recording cameras (imaging plane size, inner occular distance, focal length), the objects depth could be calculated. A Z-buffer was created, (a large matrix into which the z values for each pixel is stored). Each element in the Z-buffer matrix encodes the depth of the corresponding picture element from the video taped scene.

Several problems were encountered with this approach. Firstly, the description of the scene in the form of a Z-buffer required a lot of storage (540 x 480 elements, or approx. 300K bytes) and was very low in information content, only values for edges were calculated leaving the rest of the image with a depth value of zero, furthermore it would require significant additional processing to determine the orientation of an object from the Z-buffer. Secondly, the Z-buffer has to be hand-edited since the boundary detection heuristic, in this case a convolution, failed in some areas of the scene. (overexposed or unbalanced cameras would produce incorrect matching, or loss of the picture information to make a proper match) Also, one would have to edit the Z-buffer if they wanted to add depth levels for the object surfaces to those for the edges. Thirdly, an animation system would have to be specially modified to use the Z-buffer to correctly clip computer-generated objects with respect to the underlying real-world video scene. In effect, the animation system would have to generate the artificial scene component's depth value and compare it, pixel by pixel, with the Z-buffer of the real-world scene to determine which pixel was closer to the viewer for hidden surface removal.

INTERACTIVE CORRESPONDENCE

Development of a simple interactive computer program, for describing the real-world scene modified the AI approach to specifically address the problem of animation with this technique. The massive storage and convolutions for the correlation, matching the same point in left and right views, are eliminated. The correlation problem being 90% of the work for automated scene analysis with computational stereo. The user could then input and define selected objects within the scene that were relevant to the animation. This approach made it possible to implement the depth-from-disparity algorithms on a modest system, one that many independent artists have access to. Another advantage of the interactive system was that one could extend it from static objects to sequences where the objects move within a scene creating a travelling matt similiar to rotoscoping techniques.

One problem with the small system was how to store the video for use with the interactive program. The video image could have been digitized and separated into left and right views, then scaled back to full frame resolution. A frame could have been recorded onto the write once disc. In this case, the frame was aligned with the PLZT system, then the alternating views were recorded separately so that each view would retain full resolution. The scene was also recorded in stereo for use as a 3_D TV piece. If a large memory is available, the one multiplexed tape digitized into computer memory would afford the greatest control over the image (and manipulation of the image).

THE DEPTH FROM DISPARITY TEST

At a graphic work station, the user keys up one view of the stereo video under a conventional frame buffer and traces the objects of interest from the real-world image into the computer's memory using a digitizing tablet and puck. As the objects are traced a wire frame is drawn on the display over the image for visual feedback as it is stored in the frame buffer. Each object is given a mnemonic by the animator with which it can be referred to later for editing, modifying etc. After all the objects are traced for one view, the other view may then be displayed under the frame buffer (repeating the process for the first view, only this time the user is entering the corresponding point to the alternate view). The user enters an object name, the wire frame for the first recorded polygon, for the given object, is drawn for the first view, a line extending from the first point of the first polygon to the current cursor position is displayed. While watching the cursor on the monitor, the user moves the cursor until the corresponding point on the second view is located and determined to be a stereo match. The user then indicates to the program by pressing a control key on the puck, the screen x and y displacement of the point and enters it into the object description. From this the horizontal disparity (displacement in the x direction) is determined. With this method, the description of the scene is compact (less than 300K) and at a level of detail relevent to the animator. The program allows for the scene description to be edited and further elaborated as all the information is stored in files. The entire process can be done in steps if the scene is complicated. The program was written in PL/1 and structured on a LISP type file system.

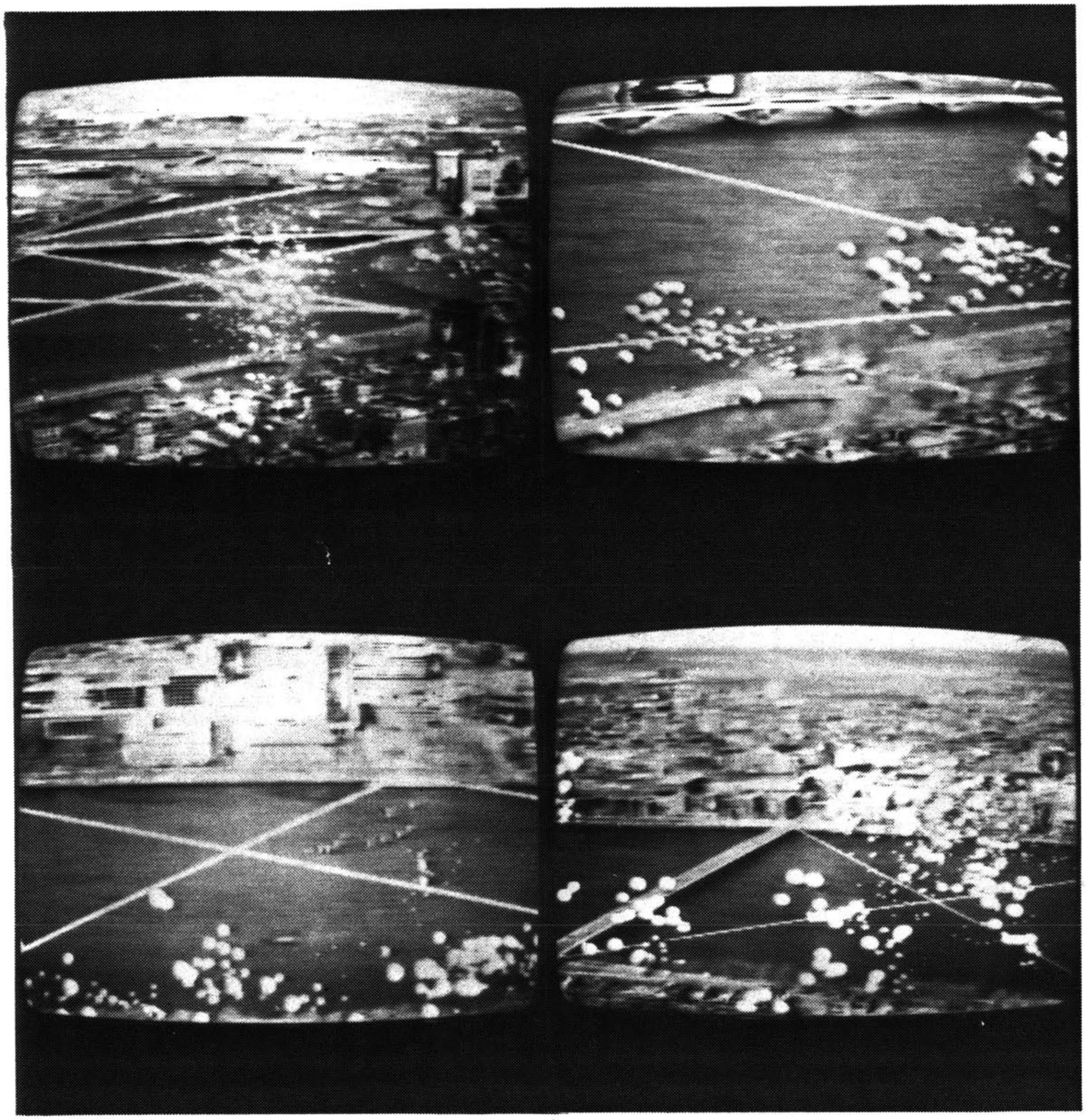
CALCULATING DEPTH FROM DISPARITY

Computational stereo is a technique for recovering three-dimensional information about a scene from two or more images of the scene taken from different points of view. Each of the different uses of stereo have different requirements that affect the design of the complete system. The stereo images may be recorded simultaneously, or at different times; the separation of the cameras can be small or large; the cameras can be parallel or at a small angle; the time of day and atmospheric conditions may be important for image quality. In this case. the conventional stereo technique was used for the project by recording two images simultaneously with laterally displaced cameras.

The disparity is inversely proportional to the depth of the point in space. A representation of the important geometrical and physical attributes of the stereo cameras, called a camera model, is used along with the disparity to create a three-dimensional coordinate system of the video scene. A function for converting the diaparity and the camera model into a depth value is called (a simple version of the Grimson algorithm) and the value of the depth is stored with the object description. A second function, which takes a depth and its screen X and Y coordinates as input and returns a corresponding world-coordinate X, Y and Z to complete the object description.

GENERATING THE ANIMATION

Once the world coordinates have been recovered, they can then be passed to the animation system. To the animation system, the objects appear as flat, two-dimensional polygons placed at various depths and are parallel to the front of the view volumn (front surface of the display). The polygons are incorporated into the list of animated objects along with the computer generated objects and given the background color value (in this case black). The computer generated objects retain their three-dimensionality and are able to be moved (animated) within the scene revealing more than one of their sides. All of the objects are placed in the coordinate system of one view (monocular) with the proper scaling, transformations etc.. When the synthetic picture is keyed over a monocular view of the scene, the video images are displayed through their 'black' computer counterparts, thus creating the illusion that the addition of the solid computer objects are being placed into the three-dimensional video 'real-world' scene.



Simulated Celebrations

Computer graphics painted over monocular videotaped image. Positioning each element in every frame by hand. Most digital effects units work in this way, some have joysticks for moving the elements around in real-time, but the artist is confined to using two-dimensional shapes that can not rotate around their z-axis.

EXPERIMENTAL VIDEOTAPES

I MARBLEHEAD

black and white (one tube NEWVICONS)
sound
PLZT field rate stereo
8 minutes

II BOSTON: CITYSCAPE

black and white (one tube NEWVICONS)
sound
PLZT field rate stereo
30 minutes

III CARDBOARD CITY

color (2/3" one tube SATICONS)
silent
PLZT field rate stereo
monocular, full frame
80 minutes

IV BOSTON: HARBOR

color (1" one tube SATICONS)
sound
PLZT field rate stereo
computer graphics overlays in stereo
5 minutes

MARBLEHEAD

This study explored object-to-ground relationships, contrast, texture and subtle stereo effects. The distance between cameras approximated normal human disparity. The cameras' height also approximated a normal viewing height of eye level while standing. The textures (water and rocks) and the contrast between the textural surfaces formed a strong aesthetic element. The sound track was recorded in stereo. In contrast to the usual nervous energy level associated with working with high tech equipment, the sound of the gulls, surf and water vehicles was extremely relaxing.

The limitations were the usual ones associated with video equipment, along with the additional constraints of long recording and composing times. Thus the recording site had a maximum distance of 200 feet from the electrical source. This prevented close-up shots for details, i.e., overlapping surfaces.

BOSTON: CITYSCAPE

Cityscape was recorded for the test at the AI lab. As the edge detection operations use light/dark area matching heuristics; the scene was composed for maximum contrast between the buildings and the sky. A soundtrack was recorded although it was not relevant to the test.

The recording was made under overcast skies. The size of the recording monitor limited the ability to align the cameras and to properly check the alternating video signal levels. The addition of a wave form monitor would have helped with regulating the voltage output from the cameras. These problems resulted in loss of information for some areas of the picture, but there was enough information to produce a depth map for most of the edges.

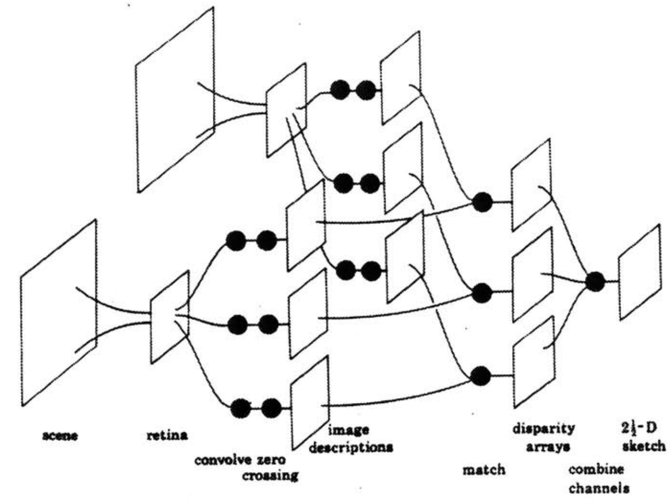
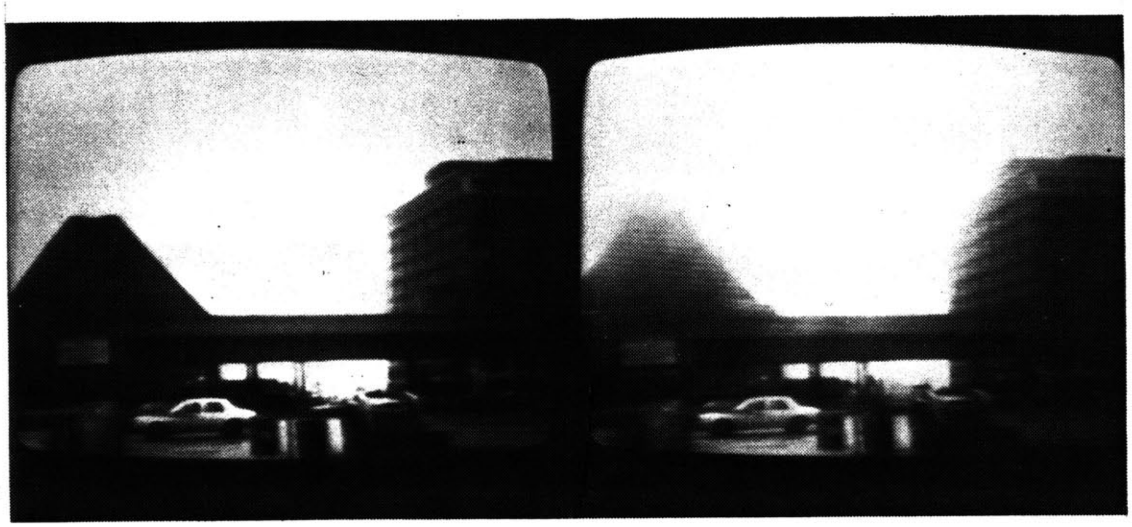


Diagram of the implementation. From the initial arrays of the scene, a retinal pair of images are extracted. These subimages reflect the current eye positions relative to the scene, and may be considered as the representation of the light intensities striking the retina. Each retinal image is convolved with a set of different-sized filters of the form $\nabla^2 G$, and the zero crossings of the output are located. For each filter size, the zero crossings descriptions are matched, based on the sign and orientation of each zero crossing point. The disparity arrays created for each filter are combined into a single disparity description, and information from the larger filters can be used to verge the eyes, thereby bringing the smaller filters into a range of operation.



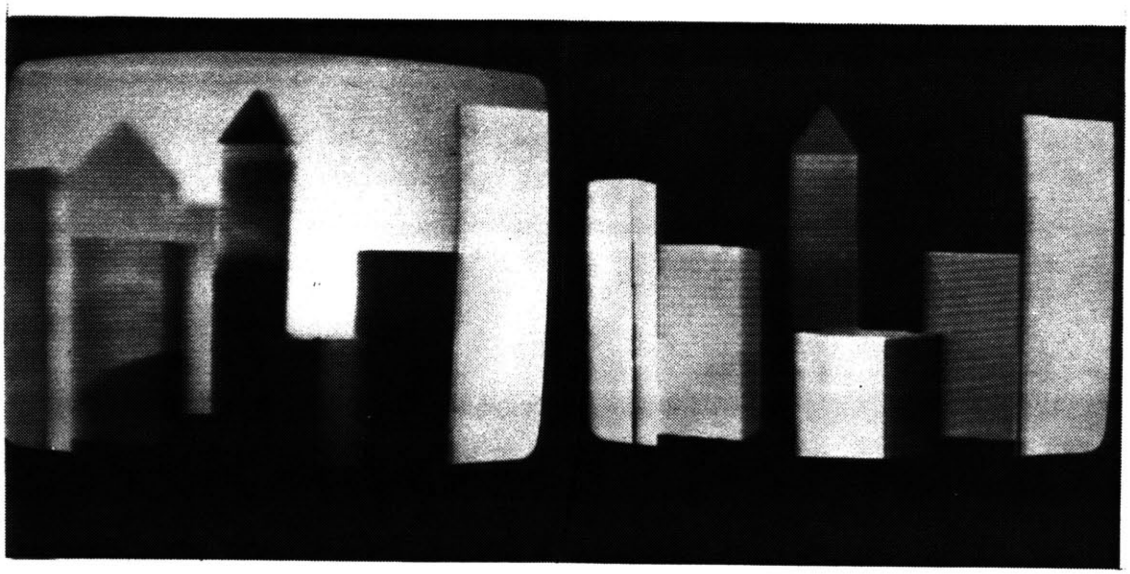
Boston Cityscape LEFT RIGHT

CARDBOARD CITY

The Cardboard City consisted of several colored boxes arranged on a turntable. The tape was recorded with strong, contrasting light on the models and white background. The turntable was rotated to several positions to obtain various viewing angles. A grid was constructed on the surface of the turntable with an origin (To) corresponding to a stationary point on the floor (Fo). The floor point (Fo), camera position, distance between the cameras and the height of the cameras were measured on a grid relative to a room coordinate (Ro). When the turntable revolved, the displacement of the objects could be determined from the angle of rotation (of turntable) by the distance between the To and Fo, the new coordinates of the objects were then given coordinates relative to the Ro.

The scene was recorded with the PLZT system (alternating field stereo) and full frame (monocular) for each camera view (left and right) so that the points for the hand-eye correlation could be discernible.

After the depth-from-disparity test, depths generated by the program were compared to the points noted in the studio.



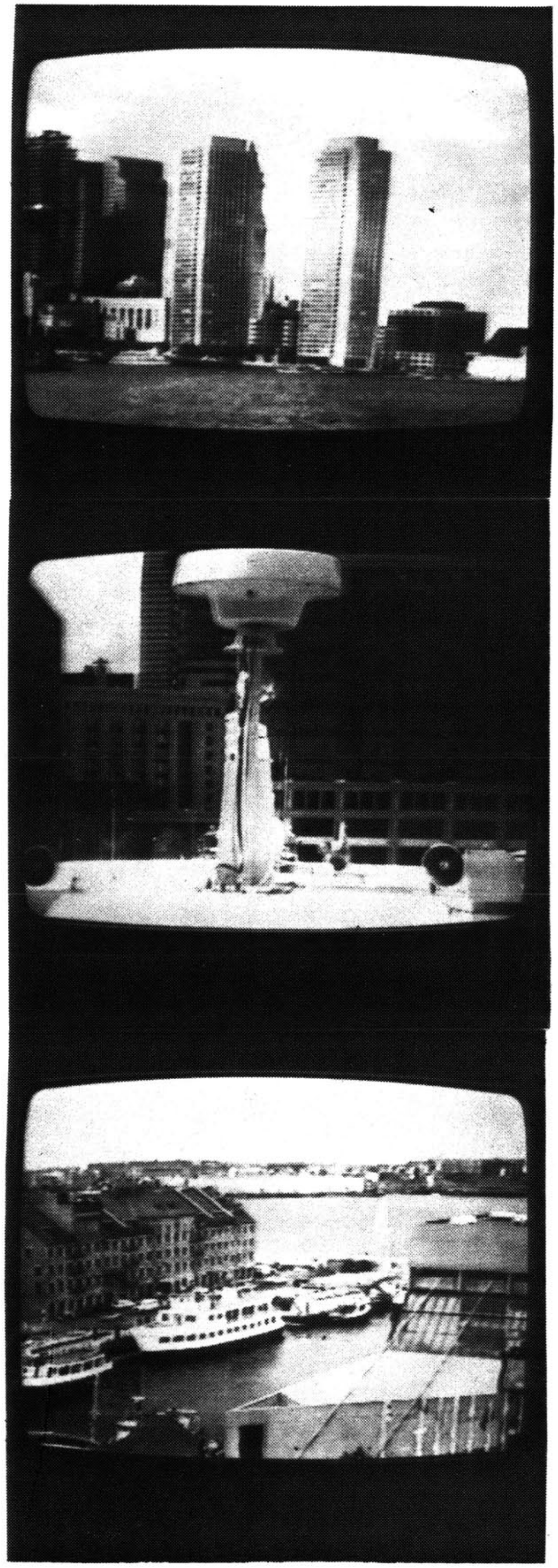
Cardboard City
videotaped image

computer generated depth-map

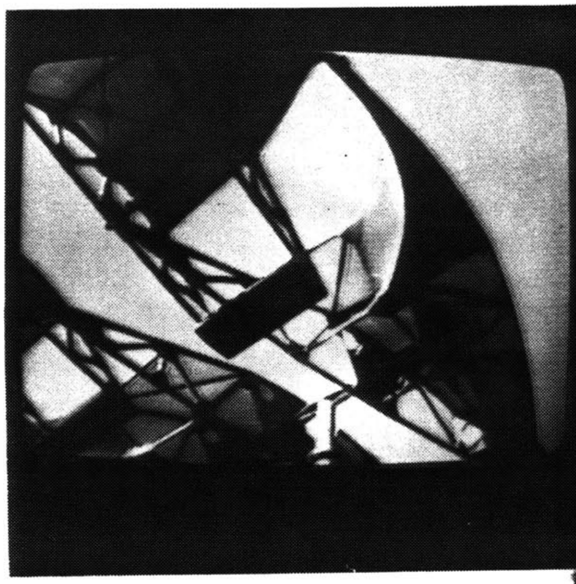
BOSTON HARBOR

Realistic scenes are combined with computer generated graphics in this stereo treatment of the Long Wharf area of Boston Harbor. The primary concern was to create a spatial exploration using the traditional film structuring techniques (long, medium, close shots; editing). The secondary concern was to present the tape in an 'environmental' situation as part of an exhibition on three-dimensional imaging techniques (Boston Museum of Science, October 1983-February 1984).

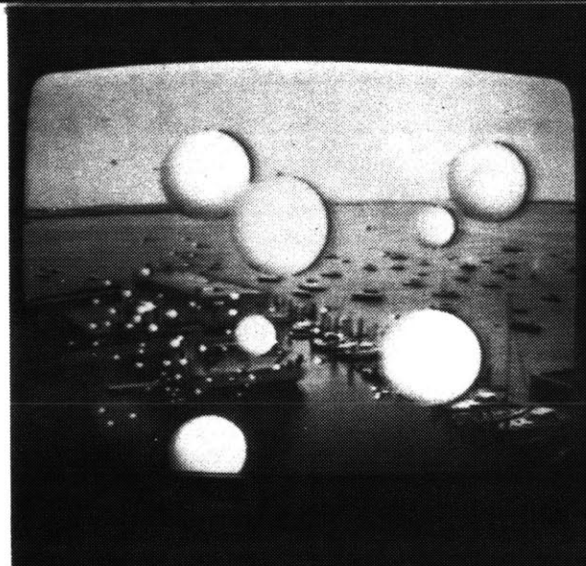
Several spatial relationships were considered: the space as created by the films' structure; the space between the frame elements (depth and virtual screen-to-object relationship); and the space of the installation.



Scenes from Boston Harbor




FORT WARREN



ACKNOWLEGDEMENTS

I would like to thank: Malakhi Simmons for not complaining about the bad food at home; John Correa and Ken Carson for helping with the technical problems; Mark Lazan and John Thompson for programming the depth-from-disparity test; Otto Piene, Director of the Center for Advanced Visual Studies, and Richard Leacock, Director of the Film/Video Section, for providing me with a place to do my work; Stephen Benton for introducing me to spatial imaging and his continuing support; Derith Glover for reading and making useful suggestions; Dorothy S., Walter B., Kate P., Sarah D., Anne O., Orlean R., Sarah G., Mark A., and the Council for the Arts at MIT for giving me hope when things looked dim; Rachele and Benjamine for accepting me into the program and continuing to believe in my talent.

Sources of Illustrations and Captions

Page

5 Electronic Light Ballet, Otto Piene. Gene Youngblood, Expanded Cinema, p 300.

6 Unmarked Interchange, Once Group, Gene Youngblood, op. cit., p 375.

6 Alwin Nikolais dancers, Gene Youngblood, op. cit., p 271.

7 Moonwalk, Theo Kameche, Amos Vogel, Film as a Subversive Art, p 187.

8 Tronhouse, Boston Globe, July 20, 1982.

10 Shadow Projection, Peter Campus, Gregory Battcock, New Artists Video, p 101.

12 Reindeer, Cave of Font-de-Gaume, France, 15000 BC, Helen Gardner, Art Through the Ages, p 41.

13 Perspective, Malton 1800, reproduced in M. H. Pirenne, Vision and Art, Handbook of Perception, p 439.

15 The First Photograph, Nicephore Niepce, 1826, Jonathan Benthal, Science and Technology in Art Today, p 27.

16 Camera Obscura, Eugene Hecht and Alfred Zajac, Optics, p 158.

18 Boat Leaving The Port, Lumiere, Gerald Mast, A Short History of the Movies, p 36.

18 Virginity, Robert Rosselini, Amos Vogel, op. cit., p 118.

19 Thaumatrope, American Cinematographer, April 1922.

20 Zoetrope, Michel Frizot, E. J. Marey 1830-1904, p 91.

20 Horse, Eadweard Muybridge, Animals in Motion, p 50.

20 Mutoscope, Erik Barnouw, The Magician and the Cinema, p 71.

20 Flip-card book, Erik Barnouw, op. cit., p 71.

22 Modern Times, Charles Chaplin, 1936.

23 The Man With The Rubber Head, Paul Hammond, The Marvellous Melies, pp 99-100.

24 Miniature and Models for Motion Picture Use, American Cinematographer, November 1922.

Illustrations continued

24 Boundary Scale for the Motion Picture Studio, American Cinematographer, June 1922.

25 Dupy Duplicator, Nora Lee, Motion Control, American Cinematographer, May 1983, p 60.

26 KING KONG, 1933 film, reproduced in Amos Vogel, op. cit., p 66.

28 Rosa Bosom, Bruce Lacey, Jasia Reichardt, Cybernetic Serendipity, p 39.

28 Mate, Bruce Lacey, Jasia Reichardt, op. cit., p 39.

28 Albert 1967, John Billingsley, Jasia Reichardt, op. cit., p 45.

28 Robot-K456, Nam June Paik, Jasia Reichardt, op. cit., p 42.

29 Scene Analysis Diagram, Ruzena Bajcsy, Three-Dimensional Scene Analysis, IEEE, 1980, p 1064.

30 Eye and Astronomer: Direct Method for Measuring Distance, R. L. Gregory, The Intelligent Eye, pp 98-99.

31-32 Binocular Video System, Alfhons Schilling, from pictures supplied to author from artist.

33 Map of U. S. with Head Mounted Display, Ivan Sutherland, Douglas Davis, Art and the Future, p 104.

35 Stereovideo System Diagram, Stephen Benton, Spatial Imaging Course Lecture Diagram, 1982.

35 PLZT Wafer Diagram, John Roese, U. S. Patents # 3821466 and 4424529.

Notes

1 Gerald Mast, A Short History of the Movies, Bobbs-Merrill Company, New York, 1971, p 177.

2 Nam June Paik, Input-Time and Output Time, Video Art, Schneider and Korot, Harcourt Brace Jovanovich, New York, 1976, p 98.

3 Peter Sorensen, Movies, Computers and the Future, American Cinematographer, January, 1983, p 78.

4 Nam June Paik, Videa 'n' Videology 1959-1973, Everson Museum of Art, Syracuse, New York, 1974, p 64.

5 James Hilton, Lost Horizon, William Morrow, U.K., 1933, p 27.

6 Amos Vogel, Film as a Subversive Art, Random House, New York, 1974, p 11.

7 Gyorgy Kepes, Language of Vision, Paul Theobald, Chicago, 1944, p 76.

8 Bearden and Holly, The Painter's Mind: A Study of the Relations of Structure and Space in Painting, Garland Publishing Inc, New York, 1981, p 217.

9 Gyorgy Kepes, op. cit., p 67.

10 Ingrid Carlbom and Joseph Paciorek, Planar Geometric Projections and Viewing Transformations, Computing Surveys, December, 1978, p 472.

11 Gyorgy Kepes, op. cit., p 109.

12 Gyorgy Kepes, op. cit., p 109.

13 Gyorgy Kepes, op. cit., p 88.

14 Gerald Mast, op. cit., p 23.

15 Paul Hammond, Marvellous Melies, St. Martin's Press, New York, 1974, p 49.

Notes continued

16 Nora Lee, Motion Control, American Cinematographer, May 1983, p 61.

17 Dan Slater, Rob King and John Gail, The Use of Computer Technology in Magicam Slave Camera System, Proceedings of the National Computer Conference, 1980, p 87.

18 Nora Lee, op. cit., p 60.

19 Jack Burnham, Beyond Modern Sculpture, Braziller, New York, 1968, p 316.

20 Jack Burnham, op. cit., p 317

21 Jack Burnham, op. cit., p 315

Bibliography

Appelbaum, Ralph, Magicam Works Magic for COSMOS, Filmmakers, Vol. 13, No. 12, October 1980.

Armes, Roy, Film and Reality, Penguin Books, U.K., 1974.

Bajcsy, Ruzena, Three-Dimensional Scene Analysis, Fifth ICPR, Vol. 2, IEEE, 1980.

Barnard, Stephen and Fischler, Martin, Computational Stereo, Computing Surveys, Vol. 14, No. 4, December 1982.

Barnouw, Erik, The Magician and the Cinema, Oxford University Press, Oxford, 1981.

Battcock, Gregory, New Artists Video, Dutton, New York, 1978.

Bearden and Holly, The Painter's Mind: A Study of the Relations of Structure and Space in Painting, Garland Publishing, Inc., New York, 1981.

Benthall, Jonathan, Science and Technology in Art Today, Praeger, New York, 1972.

Brautigam, Otto, Double Exposures of the Early Days, American Cinematographer, September 1922.

Burnham, Jack, Beyond Modern Sculpture, Braziller, New York, 1968.

Carlbom, Ingrid and Joseph Paciorek, Planar Geometric Projections and Viewing Transformations, Computing Surveys, December 1978.

Cornsweet, Tom, Visual Perception, Academic Press, New York, 1970.

Davis, Douglas, Art and the Future, Praeger, New York, 1975.

Fisher, Scott, Viewpoint Dependent Imaging: An Interactive Stereoscopic Display, unpublished MSVS Thesis, MIT Department of Architecture, 1981.

Frizot, Michel, E. J. Marey 1830-1904, Centre National D'Art Et De Culture Georges Pompidou, Musee National D'Art Moderne, Paris 1977.

Gardner, Helen, Art Through the Ages, Harcourt, Brace and World Inc., New York, 1936.

Gregory, R. L., The Intelligent Eye, Mc Graw-Hill Book Company, New York, 1970.

Grimson, Eric, A Computer Implementation of a Theory of Human Stereo Vision, Philosophical Transactions of the Royal Society of London, 1981.

Bibliography continued

Grimson, Eric, From Images to Surfaces, MIT Press, Cambridge, 1981.

Grob, Bernard, Basic Television, Principles and Servicing, Mc Graw-Hill New York, 1975.

Gulick, W. Lawrence and Robert B. Lawson, Human Stereopsis, Oxford University Press, Oxford, 1976.

Hammond, Paul, Marvellous Melies, St. Martin's Press, New York, 1975.

Hitchins, Alfred, Miniatures and Models of Motion Picture Use, American Cinematographer, November 1922.

Horn, B., Machine Vision, accepted for publishing, 1983.

Ivins, William, Three Renaissance Texts on Perspective, Metropolitan Museum of Art, New York, 1938.

Jenkins, Charles, Animated Pictures, McQueen, Washington, D. C., 1898.

Kepes, Gyorgy, Language of Vision, Paul Theobald, Chicago, 1944.

Kepes, Gyorgy, Nature and Art of Motion, Braziller, New York, 1965.

Land, Richard and Ivan Sutherland, Real-Time, Color, Stereo, Computer Displays, Applied Optics, Vol. 8, No. 3, March 1969.

Lawder, Standish, The Cubist Cinema, New York University Press, New York, 1975.

Lazan, Mark, Computer Animation Within Videospace: Depth Determination Through Computational Stereo, unpublished term paper, Computer Graphics, MIT 1983.

Lee, Norma, Motion Control, American Cinematographer, May 1983.

Lipton, Lenny, Foundations of the Stereoscopic Cinema, Van Nostrand Reinhold Company, New York, 1982.

Marr, David, Vision, W. H. Freeman and Company, San Francisco, 1982.

Mast, Gerald, A Short History of the Movies, Pegasus, New York, 1971.

Mehrdad, Azarmi, Electronic Technology in Optical Effects, American Cinematographer Manual, 1980.

Naimark, Mark, Spatial Correspondence: A Study in Environmental Media, unpublished M.S.V.S. Thesis, MIT Department of Architecture, 1979.

Okoshi, Takanori, Three-Dimensional Imaging Techniques, Academic Press, New York, 1976.

Bibliography continued

Paik, Nam June, Videa 'n' Videology 1959-1973, Everson Museum of Art, Syracuse, New York, 1974.

Patterson, Richard, Ultimate, American Cinematographer, October 1982.

Pirenne, M. H., Vision and Art, Handbook of Perception, Volume V, Seeing, editors E. Carterette and M. Friedman, Academic Press, New York, 1975.

Pirenne, M. H., Optics, Painting and Photography, Cambridge University Press, Cambridge, 1970.

Potmesil and Chakravary, Synthetic Image Generation with a Lens and Aperture Camera Model, Transactions on Graphics, ACM, April 1982.

Reichardt, Jasia, Cybernetic Serendipity: The Computer and the Arts, Studio International, London, 1968.

Richard, Paul, Robot Manipulators, MIT Press, Cambridge, 1981.

Rickey, George, Constructivism, Origins and Evolution, Braziller, New York, 1967.

Roese, John, Liquid Crystal Stereoscopic Television System, United States Patent #3821466, June 28, 1974.

Roese et al., Remotely Triggered Portable Stereoscopic Viewer System, United States Patent #4424529, January 3, 1984.

Roese, John, Stereoscopic Computer Graphics Using PLZT Electro-Optic Ceramics, Proceedings of SID, 1978.

Rotzler, Willy, Constructive Concepts: A History of Constructive Art From Cubism to the Present, ABC Edition, Zurich, 1977.

Runcie, Osborne, Stereoscopy and Its Application to Cinematography, American Cinematographer, January 1922.

Russett, Robert and Cecile Starr, Experimental Animation, Van Nostrand Reinhold, New York, 1976.

Ryan, Paul, Three-Dimensional Cinematography for Magic Journeys, American Cinematographer, February 1983.

Schneider, Ira, and Beryl Korot, Video Art: An Anthology, Harcourt Brace Jovanovich, New York, 1976.

Slater, Dan and Rob King and John Gale, The Use of Computer Technology in Magicam Slave Camera Systems, Proceedings, National Computer Conference, 1980.

Bibliography continued

Sobel, Irwin, On Calibrating Computer Controlled Cameras for Perceiving Three-Dimenaional Scenes, AI 5, 1974.

Sorensen, Peter, Movies, Computers and the Future, American Cinematographer, January 1983.

Sorensen, Peter, Tronic Imagery, Byte, November 1982.

The First Moving Picture Machine, American Cinematographer, April 1922.

Thompson, John, Computer Animation Within Videospace: Real-World Object Description for Computational Stereo Analysis of Static Scenes, unpublished term paper, Computer Graphics, MIT, 1983.

Tuchman, Maurice, A Report on the Art and Technology Program of the Los Angeles Country Museum of Art 1967-1971, LACMA, 1971.

Vogel, Amos, Film as a Subversive Art, Random House, New York, 1974.

Whitney, John, Motion Control: An Overview, American Cinematographer, December 1981.

Williams, Alan, Convergence Techniques for Three-Dimensional Filming, Americam Cinematographer, March and April 1984.

Wingler, Hans, The Bauhaus, MIT Press, Cambridge, 1981.

Wolfe, Tom, The Right Stuff, Farrar, Straus and Giroux Inc., New York, 1979.

Yakimovshy, Yoram, A System for Extracting Three-Dimensional Measurements from a Stereo Pair of TV Cameras, Computer Graphics and Image Processing 7, 1978.

Yakimosky, Yoram, Boundary and Object Detection in Real-World Images, Journal of the Association for Computing Machinery, Vol. 23, No. 4, October 1976.

Youngblood, Gene, Expanded Cinema, Dutton and Company, New York, 1970.